DOCUMENT RESUME

ED 211 559 TM 810 559

AUTHOR Bulcock, J. W.
TITLE Normalization Ridge Regression in Practice I: Comparisons Between Ordinary Least Squares, Ridge Regression and Normalization Ridge Regression.
INSTITUTION Uppsala Univ. (Sweden). Dept. of Psychology.
SPONS AGENCY Natural Sciences and Engineering Research Council, Ottawa (Ontario).; Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).; Swedish Inst., Stockholm.; Uppsala Univ. (Sweden). Inst. of Theology.
PUB DATE May 81
GRANT A-3960
NOTE 35p.; For related document, see TM 810 560. Lecture given at University of Uppsala (Sweden, May 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Covariance; *Least Squares Statistics; *Mathematical Models; *Predictor Variables; Regression (Statistics); *Research Methodology; Research Problems; Social Science Research
IDENTIFIERS *Normalization Regression Estimation; *Ridge Regression Analysis; Stochastic Analysis

ABSTRACT

The problem of model estimation when the data are collinear was examined. Though the ridge regression (RR) outperforms ordinary least squares (OLS) regression in the presence of acute multicollinearity, it is not a problem free technique for reducing the variance of the estimates. It is a stochastic procedure when it should be nonstochastic and it does not satisfy the boundary condition. It is argued that the dilemmas of the RR approach stem from use of the minimum mean square error criterion. An alternative, called the variance normalization criterion is proposed. In theory this method overcomes the dilemmas of RR while at the same time preserving all the advantages of RR over OLS. On the basis of nine performance indices, it is shown that when applied to test data normalization ridge regression performs satisfactorily. Monte Carlo experiments are necessary to shed more light on the problem of selecting the most satisfactory optimal k-value for coping with the multicollinearity problem. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED211559

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

NORMALIZATION RIDGE REGRESSION IN PRACTICE I: COMPARISONS BETWEEN ORDINARY LEAST SQUARES, RIDGE REGRESSION AND NORMALIZATION RIDGE REGRESSION*

J.W. Bulcock
University of Alberta
and
Memorial University of Newfoundland

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. W. Bulcock
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 555

* This is the text of the first of two lectures given in the Department of the Psychology of Religion, University of Uppsala, Sweden, May, 1981. The lectures were supported by a Visiting Professor appointment sponsored by the Swedish Institute and the Institute of Theology, University of Uppsala. The research was partly supported by Grant No. A3960 from the Natural Sciences and Engineering Research Council of Canada, and partly by a Leave Fellowship awarded by the Social Sciences and Humanities Research Council of Canada. Generous assistance toward computing costs was provided by the Department of Educational Foundations, University of Alberta, Edmonton, Canada.

ABSTRACT

Though ridge regression (RR) outperforms ordinary least squares (OLS) regression in the presence of acute multicollinearity, it is not a problem free technique for reducing the variance of the estimates. It is a stochastic procedure when it should be non-stochastic and it does not satisfy the boundary condition. It is argued that the dilemmas of the RR approach stem from use of the minimum mean square error criterion. An alternative, called the variance normalization criterion. is proposed which in theory overcomes the dilemmas of RR while at the same time preserving all the advantages of RR over OLS. On the basis of nine performance indices it is shown that when applied to test data normalization ridge regression performs satisfactorily.

This paper is about the problem of model estimation in the social sciences when the data is ill-conditioned, or where the data consists of conceptually distinct predictor variables which are highly intercorrelated.

In part one the problem is identified.

In part two the problem is illustrated by reference to a simple five factor model of employee satisfaction.

In part three several standard solutions are examined.

In part four selected statistical solutions are applied to the estimation of the five factor employee satisfaction model.

In part five the results of the part four estimates are evaluated in terms of performance indices, and a solution to the problem is proposed.

PART ONE: THE PROBLEM

When the researcher tries to estimate social science models under conditions of multicollinearity, standard estimating procedures such as ordinary least squares regression (OLS), two-stage least squares (2SLS), or full information

maximum likelihood (FIML) frequently yield unreliable parameter estimates. Multicollinearity refers to the presence of highly intercorrelated predictor variables in structural equation models, and its effect is to invalidate some of the basic assumptions underlying their mathematical estimation unless these mathematical methods are modified. It is not surprising, then, that it is considered "one of the most ubiquitous, significant and difficult problems in applied econometrics (Intrilligator, 1978, p. 152) and referred to by social modelers as "the familiar curse" (Wheeler, 1980, p. 14).

The fact is that the presence of high correlation, nonorthogonality, or approximate linear relationships in a set of predictors is not that uncommon for data in social science research. The situation is often difficult to detect, and until recently even more difficult to remedy. It results in large standard errors so that estimates of the true model parameters become unstable; the parameter estimates vary considerably from sample to sample; often there will be drastic changes in the estimates after only minor data revision; conflicting conclusions will be reached from the usual tests of significance; and because some estimates are reduced in significance while others may be excessively large and even of the incorrect sign, a good fit need not indicate good forecasting power. There are numerous studies of the multicollinearity problem which verify these pronouncements (Althauser, 1971; Blalock, 1963, 1964; Christ, 1966; Farrar and Glauber, 1967; Gordon, 1968; Intrilligator, 1978; and Johnston, 1972).

One of the best practitioners of quantitative social science has commented on the problems besetting today's social modeler, and his remarks are paraphrased as follows:

> ... an uneasy feeling about the present state of social modeling has been growing in some of us who have watched its unprecedented development over the last three decades. This concern seems to be shared even by those who are themselves contributing successfully to the present boom. They play the game with professional skill but have serious doubts about its rules.... The uneasiness ... is caused not by the irrelevance of the practical problems to which social scientists address their efforts, but rather by the palpable inadequacy of the scientific means with which they try to solve them ... the consistently indifferent performance in practical applications is in fact a symptom of a fundamental imbalance in the present state of social modeling. The weak and all too slowly growing empirical foundation clearly cannot support the proliferating structure of pure, or should I say, speculative social science theory.

(A paraphrase of the opening remarks of Wassily Leontif in his Presidential Address to members of the American Economic Society, 1971).

PART TWO: AN ILLUSTRATION

The problem of the estimation of collinear data sets is illustrated by reference to Bertram Price's (1977) five factor model of employee satisfaction with their supervisors. The dependent variable, y, was the employee's assessment of the overall job being done by his or her supervisor. The predictors, selected by management, were employee ratings of how well the supervisor (i) handled employee complaints (X_1), (ii) provided opportunities to learn new things about the job (X_2), (iii) gave raises based on performance (X_3), (iv) provided criticism of people who did a poor job (X_4), and (v) applied company policy fairly (X_5). A simple conceptual model showing the direction of the hypothesized relationships is presented in Figure 1.

Figure 1 about here

Each predictor was a dichotomous index. The data on 30 departments with an average of 35 employees in each was aggregated to the department level; hence, consisted of 30 observations. The matrix of zero-order correlations is presented in Table 1. Note that several large interfactor correlations are reflected in the X'X eigenvalues. Thus, the expected squared distance of the least squares coefficient estimate $\hat{\beta}$ from β is $\sigma^2\Sigma\lambda_i^{-1}/p = 738.7\sigma^2/5$ or more than 140 times what it should be for an orthogonal system.

Table 1 about here

The least squares model was of the general form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \qquad (1)$$

And the least squares parameter estimates (standardized partial regression coefficients) were as follows:

$$y = \underset{(3.271)}{2.492*X_1} + \underset{(2.696)}{1.537*X_2} + \underset{(2.316)}{2.370*X_3} - \underset{(2.373)}{.405*X_4} - 4.720*X_5 + .484 \qquad (2)$$

where the t-values are in parentheses below the standardized beta coefficients, and where any t-value GT 2.0 is significantly different from zero at the .05 level.

There are three obvious signs that the Price data might be ill-conditioned. First, the correlations (X'X) among the predictor variables in Figure 1 are high. Second, this is confirmed by an examination of the eigenvalues which were referred to above. Third, most of the standardized betas are greater than unity; hence, virtually impossible to interpret.

Examination of the inverse of the correlation matrix $(X'X)^{-1}$ or the variance inflation factor (VIF) matrix in Table 2 confirms these suspicions that the data are ill-conditioned. The VIF or diagonal elements of the inverse of the simple correlation matrix, are measures of the collective impact of the interdependency of the explanatory variables on the variance of the regression coefficient of each term. Each

VIF, then provides an indication of the number of times the variance of each estimate ($\hat{\beta}_i$) has been inflated; where one stands for no collinearity and infinity for perfect collinearity (Marquardt and Snee, 1975, p.4). Thus, the maximum VIF (Vmax) is often considered the best single indicator of multicollinearity. As shown in Table 2 the Vmax of the Price predictor variable matrix is 421.8. The degree of multicollinearity can, however, be transformed into a scale of zero to one by defining the degree of multicollinearity "D" as:

$$D = Dmax = \frac{2}{\pi}\tan^{-1}(Vmax - 1) \qquad (3)$$

where $\tan^{-1}(Vmax - 1)$ is measured in radians.

In the case of Table 2 Dmax = .9985. It should be noted, however, that the maximum effect of multicollinearity is manifest in factor space when the model is expressed in its canonical form. Thus, it may be preferred to measure the absolute degree of multicollinearity in factor space for which $Vmax = 1/\lambda min$ and

$$Dmax = Dc = \frac{2}{\pi}\tan^{-1}(1/\lambda min - 1) \qquad (4)$$

where λmin is the smallest eigenvalue.

Note that the difference between Dc and Dmax is that Dc is an absolute measure of the degree of multicollinearity; and, therefore, a fixed value for any data set. Dmax in contrast is a relative measure in that its value changes with the data handling technique; that is, its value depends on the orientation of the axes of the variable vectors.

In the present example Dc = .999 on a scale of zero to one. Positive Dmax indicates inflated variance while negative Dmax indicates deflated or over-suppressed variance. There is little doubt from Vmax, Dmax, or Dc values that the degree of multicollinearity in the data set is excessive.

Table 2 about here

PART THREE: STANDARD SOLUTIONS

In the presence of acute multicollinearity, then, OLS regression estimates are frequently both erroneous and uninterpretable. What can be done? The total variance of regression coefficients may be expressed as (Hoerl and Kennard, 1970a):

$$\Sigma \mathrm{Var}(\hat{\beta}) = \hat{\sigma}^2 \mathrm{tr}(X'X)^{-1}$$

$$= \frac{\varepsilon'\varepsilon'}{n-p} \mathrm{tr}(X'X)^{-1} \qquad (5)$$

where Σ represents the summation from i=1 to i=p, the number of predictors; where β is the (k x 1) vector of estimated regression coefficients; ε is the random error of y in the regression model ($y = X\beta + \varepsilon$); n is the sample size; and $\mathrm{tr}(X'X)^{-1}$ is the trace of the inverse of the correlation matrix (X'X).

From Eq. (5) it is obvious that the total variance depends on three independent factors: $\varepsilon'\varepsilon$, n-p, and $\mathrm{tr}(X'X)^{-1}$. Since in any statistical procedure it is desirable to have the variance as low as possible, standard solutions to variance inflation caused by multicollinearity depend on the location of these factors followed by efforts to eliminate them if possible.

The first source of inflated variance is $\varepsilon'\varepsilon$, the random error. This is the random portion of the dependent variable, y, which generally consists of two parts; that due to measurement error ε_m, and that due to stochastic

error ε_s. Stochastic error is also attributable to two components; namely, the incompleteness of the designed model, and some unknown irreproducible fluctuation (Wannacott and Wannacott, 1969, p. 17). These two errors are assumed to be uncorrelated and to have normal distribution with zero mean. Thus, the sum of the squares error may be expressed as:

$$\varepsilon'\varepsilon = \varepsilon_m'\varepsilon_m + \varepsilon_s'\varepsilon_s \qquad (6)$$

And this error may be reduced, though not eliminated, by improving the precision of measurement, and by improving the design of the model; that is, by improved model specification.

The second source of inflation is the sample size n, or more specifically, the degree of freedom of the sum of squares error, n - p. Thus, the total variance may be reduced but not eliminated, by increasing the sample size.

But the most important source of variance inflation is that due to the effect of multicollinearity which enters through $tr(X'X)^{-1}$; that is, the trace of the inverse of the predictor variable matrix, or the sum of the variance inflation factor of each variable due to its interrelationship with the rest of the explanatory variables. This third source of variance is statistical; thus, a statistical solution is called for.

1. Statistical Solution #1: The Construction of Composite Variables

In recursive models one of the commonest methods of

coping with multicollinearity is through model simplification. This may be achieved in two ways. First, because multicollinearity indicates that some independent variables convey little information over those of other variables, one way of simplifying or scaling down a model is to drop variables. Second, a group of variables may be aggregated or averaged. A formal method of averaging may be conducted using principal components analysis or confirmatory factor analysis. In these instances the resultant composites or latent variables are conceived as reflective of their manifest indicators. See Figure 2a. If the observed variables are in a productive relationship to the latent variable a canonical model may be called for. See Figure 2b for a hypothetical example.

Figure 2 about here

Where a statistical solution to multicollinearity is called for standard packages with the capacity for estimating models with latent variables may be sufficient for coping with multicollinearity in the recursive model case. This is because the collinearity may be contained within the latent variables. Nevertheless, there are three problems with latent variable path model approaches. Though a first principal component will pick up the major portion of the variances of the manifest variables, it need not be the one which is the most strongly related to the dependent variable.

Second, the first principal component may have no theoretical meaning; hence, there may be a lack of congruence between the concepts of the theoretical model and the variables in the measurement or auxiliary model. Third, conceptually distinct latent variables may still be highly intercorrelated.

2. Statistical Solution #2: Ridge Regression

Ridge regression (RR) developed by Hoerl and Kennard (1970a, 1970b) is becoming generally accepted in the natural sciences and engineering as the most promising alternative to OLS for coping with the multicollinearity problem. See for example, Vinod's survey paper (1978). The ridge solution is based on the minimum total mean square error criterion (MSE) as opposed to the minimum sum of squares criterion of OLS. From a geometrical point of view MSE represents the expected squared distance between an estimator, $\hat{\beta}$, and the true vector, β. Since the best estimator is the one that minimizes this distance the MSE criterion would seem to be a viable alternative to the OLS criterion. The actual ridge technique involves augmenting the diagonal of the normal equation matrix prior to inversion with a small positive quantity k -- the biasing parameter. The argument underlying the functional formulation of RR is that in the presence of multicollinearity an estimator with a small bias and small variance is to be preferred to one that has no bias but a large variance.

Because in OLS the bias is zero, when the predictors

are highly correlated the variance usually becomes large. The effect of the ridge adjustment which involves the introduction of some bias, is to decrease the variance. In the preferred case, introduction of the ridge adjustment will exchange a large reduction in variance for a small bias, thereby reducing the MSE. When this occurs the ridge coefficients will be closer to their true values than OLS coefficients.

In simple RR (as opposed to generalized RR) the estimated regression coefficient vector of a linear model, $y = X\beta + \varepsilon$, is

$$\hat{\beta}^* = (X'X + kI)^{-1}X'y \qquad (7)$$

as opposed to that of OLS regression,

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (8)$$

where for convenience the X's and y are standardized so that X'X gives the correlation matrix among the independent variables and X'y gives the correlation vector between the dependent and independent variables; and where k is the biasing parameter which produces minimum MSE (Hoerl and Kennard, 1970a).

Several RR techniques have been developed -- about fifteen at the last count. All use the total MSE function as the criterion for measuring the "goodness" of an estimator; that is, all use this criterion in an effort to shrink the least squares estimator toward the origin in order to yield estimates with smaller $\hat{\beta}^*$ to β distances than OLS estimators. Vinod (1978) notes that in some sixteen Monte

Carlo simulation experiments in which the estimated regression coefficients using RR and OLS are compared in terms of MSE, "the superiority of RR over OLS is almost always noted." In the best known of these studies, Dempster (1973) reported results which strongly favoured RR over OLS when multicollinearity was present. Thus, disagreement centres not about RR vs. OLS; but rather, which one of several RR methods generates the optimal biasing parameter.

The RR approach goes much of the distance toward coping with the harmful effects of multicollinearity among conceptually distinct predictor variables. It produces smaller sampling variance than OLS (Bulcock et al, 1980, pp. 36, 44). Because RR produces smaller sampling variance the RR estimate is less sensitive to sampling fluctuation. Third, because stochastic error is a source of the total variance of the regression estimates (cf. Eq. 3) and because RR reduces the variance of the regression estimates, the ridge solution is less sensitive to model misspecification than OLS solutions. Fourth, since the higher the variance the higher the linear dependence of the estimates -- that is, the more likely the beta coefficients will be correlated -- any procedure reducing the variance will generate more accurate estimates. Fifth, the ridge technique generates a shorter coefficient vector (for proof see Bulcock et al. 1980, pp. 32-34). Sixth, when a data set is ill-conditioned, OLS models do not provide accurate prediction -- not even if the R-square is high. But even thoughRR produces smaller

R-square than OLS for recursive models (for proof see Bulcock et al. 1980, pp. 43-44) the procedure always produces a more accurate prediction equation if the biasing parameter is not too large (for proof see Bulcock et al. 1980, pp. 44-45).

These are compelling reasons for using RR in preference to OLS for coping with, but not necessarily solving, the multicollinearity problem. Yet, despite the important advantages of RR three of its sixteen properties are problematic (properties 8, 15 and 16 in Bulcock et al. 1980, pp. 40-41, 47-50), which give rise to four dilemmas. The optimal k can only be calculated if the true β and σ^2 are known -- but these are unknown population parameters so that k has to be estimated from sample data. Second, for this reason the procedure is stochastic when it should be nonstochastic. Thus, RR does not give a single automatic solution to the estimation problem, but rather a family of solutions. In the third place, the RR technique does not satisfy the boundary condition and, therefore, for othogonal data produces different estimates from those of OLS (Hoerl and Kennard, 1970a). Fourth, RR is a multifunctional technique. In suppressing the VIF it is suppressing not only the variance inflation due to multicollinearity but also the variance inflation attributable to other causes of inflation; namely, measurement error, model misspecification and small sample size.

3. Statistical Solution #3: Normalization Ridge Regression

The outstanding question, then, is whether an estimating procedure can be developed which retains the advantages of the standard RR procedure while overcoming its inherent limitations. The four dilemmas of RR stem from the use of the MSE criterion. But Marquardt (1970) and Marquardt and Snee (1975) have suggested a value for the biasing parameter, k, for which the minimum VIF is "between one and ten and closer to one." It has been shown that in order to perform the task of reducing variance inflation due to multicollinearity, and only multicollinearity,, the VIFs should be normalized such that each is equal to unity.

The lowest permissible variance in an equation is $p\hat{\sigma}^2$, which is the value of the best condition; namely,

$$\Sigma \text{Var}\,(\hat{\beta}^*) = p\hat{\sigma}^2$$

$$\text{but } \sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = p\sigma^2$$

$$\text{therefore } \frac{1}{p} \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = 1 \qquad (9)$$

In other words, in order to reduce variance inflation due to multicollinearity one should normalize the VIF so that it is equal to one. The value of k which satisfies this condition is the k which satisfies equation 9.

PART FOUR: APPLICATION OF RR AND NR SOLUTIONS

Application of the normalization ridge regression (NR) method and six different RR methods to the estimation of Price's five factor employee satisfaction model generate the results presented in equations 16 through 22. The formula for calculating the NR biasing parameter is given as Eq. (9). The six RR formulas are:

(i) Kasarda and Shih (1977)

$$MSE = \sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} + k\Sigma \frac{\hat{\sigma}_i^2}{(\lambda_i + k)^2} = \text{minimum} \qquad (10)$$

(ii) Hocking, Speed and Lynn (1976)

$$k = \hat{\sigma}^2 \frac{\Sigma \lambda_i^2 \alpha_i^2}{\Sigma \lambda_i^2 \hat{\alpha}_i^4} \qquad (11)$$

(iii) Lawless and Wang (1976)

$$k = \frac{p\hat{\sigma}^2}{\Sigma \lambda_i \hat{\alpha}_i^2} \qquad (12)$$

(iv) Hoerl, Kennard and Baldwin (1975)

$$k = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_i^2} \qquad (13)$$

(v) Hoerl and Kennard's Iterative (1976)

$$k_0 = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_i^2} \qquad (14)$$

. . . . . .

$$k_t = \frac{p\hat{\sigma}^2}{\Sigma \hat{\alpha}_{it}^{*2}}, \quad \hat{\alpha}_{it}^* = \frac{\lambda_i \hat{\alpha}_i}{(\lambda_i + k_t)}$$

with tolerance level

$$\frac{k_t + 1 - k_t}{k_t} \leq \delta$$

where $\delta = 20T^{-1.30}$

$T = tr(X'X)^{-1}/p$

(vi) Vinod (1976)

$$ISRM = \Sigma((p\lambda_i/(\lambda_i + k)^2 S)-1)^2 = \text{minimum} \qquad (15)$$

where $S = \Sigma\lambda_i/(\lambda_i + k)^2$

The estimates for the different ridge methods are as follows:

(i) Normalization:

$$y = \underset{(5.204)}{.569}*COMPLAINTS + \underset{(1.641)}{.181}*LEARN - \underset{(.204)}{.022}*RAISE + \underset{(.119)}{.011}*CRITICISM + \underset{(2.606)}{.128}*POLICY + .603 \qquad (16)$$

(ii) Kasarda and Shih:

$$y = \underset{(3.695)}{1.775}*COMPLAINTS + \underset{(2.751)}{1.008}*LEARN + \underset{(2.197)}{1.403}*RAISE - \underset{(1.579)}{.237}*CRITICISM - \underset{(2.306)}{2.820}*POLICY + .521 \qquad (17)$$

(iii) Hocking, Speed and Lynn:

$$y = \underset{(5.201)}{.627}*COMPLAINTS + \underset{(1.657)}{.200}*LEARN - \underset{(.156)}{.091}*RAISE + \underset{(.099)}{.010}*CRITICISM + \underset{(.975)}{.067}*POLICY + .581 \qquad (18)$$

(iv) Lawless and Wang:

$$y = \underset{(5.193)}{.162}*COMPLAINTS + \underset{(1.642)}{.194}*LEARN - \underset{(.188)}{.022}*RAISE$$

$$\underset{(.108)}{.010*CRITICISM} + \underset{(1.386)}{.086*POLICY} + .586 \qquad (19)$$

(v) Hoerl, Kennard and Baldwin:

$$y = \underset{(3.802)}{1.663*COMPLAINTS} + \underset{(2.757)}{.926*LEARN} + \underset{(2.162)}{1.252*RAISE}$$

$$- \underset{(1.148)}{.211*CRITICISM} - \underset{(2.287)}{2.525*POLICY} + .525 \qquad (20)$$

(vi) Hoerl and Kennard's Iterative:

$$y = \underset{(5.281)}{.521*COMPLAINTS} + \underset{(1.709)}{.171*LEARN} - \underset{(.102)}{.010*RAISE}$$

$$+ \underset{(.116)}{.010*CRITICISM} + \underset{(3.855)}{.156*POLICY} + .627 \qquad (21)$$

(vii) Vinod (first value):

$$y = \underset{(5.273)}{.691*COMPLAINTS} + \underset{(1.786)}{.231*LEARN} + \underset{(.108)}{.051*RAISE}$$

$$+ \underset{(.035)}{.003*CRITICISM} - \underset{(.358)}{.038*POLICY} + .565 \qquad (22)$$

(viii) Vinod (second value):

$$y = \underset{(6.678)}{.275*COMPLAINTS} + \underset{(3.237)}{.138*LEARN} + \underset{(2.171)}{.082*RAISE}$$

$$+ \underset{(.254)}{.012*CRITICISM} + \underset{(7.090)}{.170*POLICY} + .794 \qquad (23)$$

These estimates lack consistency. The Kasarda-Shih and Hoerl-Kennard-Baldwin coefficients are in several instances greater than unity; hence, uninterpretable. Vinod's index of stability method generated more than one minimum point, though in theory there is only one optimum k that gives

minimum MSE. The ridge estimates of the remaining methods differed among themselves. Thus, although according to the normalization and the Hoerl-Kennard iterative methods the COMPLAINTS and POLICY parameters were significant, according to the Hocking-Speed-Lynn and Lawless-Wang methods only the COMPLAINTS factor was significant. In view of such inconsistency, which method is to be preferred? About the only firm conclusion is that the RR results (Eqs. 16-23) do not support the OLS results (Eq. 2).

In terms of performance, then, there is little to be concluded from the estimates of the methods themselves. One has to make comparisons between performance indices in order to choose between methods.

PART FIVE: COMPARING PERFORMANCE INDICES

In RR the bias depends on unknown population coefficients (Hoerl and Kennard, 1970a), therefore, it is not possible to use a single measure such as MSE which involves the bias introduced by the ridge procedure as a performance measure. RR analysts recognize this and have developed multiple performance indices accordingly. Analysts such as Hoerl and Kennard (1970a) and Marquardt and Snee (1975) compare the different biasing parameters, the k's; the inflated sum of squares, ϕ's; and the maximum variance inflation factors, Vmax. McCabe (1978) and Obenchain (1977) have used the α-acceptability index. Above, use of Dmax was proposed. In addition four new performance measures will be used; a sensitivity ratio, the relative reduction in variance, the average percentage error of estimated variance which have been described elsewhere (Bulcock et al. 1980, pp. 73-74), and R-square.

In general for any RR estimation method it is desirable: (i) to reduce the Vmax to a value close to unity; (ii) to generate the maximum reduction in the relative degree of multicollinearity (Dmax) without introducing too large a bias, k; (iii) to minimize the sampling variance or E.S.Ratio; (iv) to reduce the total variance of the OLS estimate (RVAR) inflated by collinearity; (v) to produce acceptably low inflation in the residual sum of squares (ϕ_0); and to generate the lowest measure of the average percentage error

of estimated variance (Av.%Er.VAR).

The results of the estimation of the five factor employee satisfaction model in terms of performance indices are presented in Table 3, and in Table 4 the data of Table 3 are ranked for each performance index in order to simplify interpretation.

Table 3 and Table 4 about here

Obviously, the variance of the OLS estimates is unacceptably high. See column 8 in Table 3 where the Av.%ErVAR = 12590.6 percent. With the possible exception of the Hoerl-Kennard-Baldwin and Kasarda-Shih results the RR methods significantly reduced this excessive variance. In the Vinod #2 case, however, the bias introduced was excessive. Thus, three methods were problematic. The Vinod method was inconsitent because it generated more than one optimal k. Kasarda and Shih's method and the Hoerl-Kennard-Baldwin method generated conservative results.

The remaining methods -- NR, Hocking-Speed-Lynn, Lawless-Wang, and the Hoerl-Kennard iterative were all acceptable in performance terms. Though the Hoerl-Kennard iterative method had an edge over the other three methods the differences between all four were marginal. It is worth noting that the NR method was probably marginally better than either the Hocking-Speed-Lynn or Lawless-Wang methods.

PART SIX: CONCLUDING REMARKS

1. In this paper the problem of model estimation when the data are ill-conditioned or collinear was examined. The task called for the making of comparisons between three estimating techniques; namely, OLS, NR, and a family of methods in the RR tradition.

2. It was pointed out that all standard structural equation estimating methods are sensitive to the problem of multicollinearity. Standard ways of coping with the multicollinearity problem include improving the precision of measurement, improving the specification of the model, increasing the sample size, and scaling down a model by dropping redundant variables or by aggregating variables through principal component analysis, factor analysis or canonical correlation analysis. Where there is lots of prior information about the model available, these methods of coping with the muticollinearity problem may be sufficient and especially in the case of recursive models.

3. Where these methods are not adequate it was argued that the analyst might wish to consider using an estimating technique which generates regression coefficients with smaller variance at the expense of introducing a small bias.

4. This is effectively what Hoerl and Kennard, the originators of RR, have accomplished. An examination of the properties of RR showed that as a collinearity reduction technique it copes satisfactorily with the problem.

Nevertheless, RR is not problem-free. Because it is a stochastic procedure when it should be nonstochastic it does not yield a single, unambiguous solution; rather, a whole family of solutions are possible. Thus, though numerous Monte Carlo experiments have demonstrated convincingly the overall superiority of RR compared to OLS in the presence of multicollinearity, there is wide disagreement about the optimum RR method.

5. The question posed, then, was whether a procedure could be devised which retains the advantages of ridge regression while at the same time coping satisfactorily with its disadvantages. The use of a new estimating criterion called the variance normalization criterion was proposed, and an effort was made to show that the use of this criterion enables the analyst to go some of the distance toward answering this question.

6. The problems of multicollinearity were illustrated through estimation of Bertram Price's (1977) data set. On the basis of nine performance indices it was shown that NR generated estimates which were as acceptable as three other sets of RR estimates. Three RR methods were found to be inferior to these.

7. Monte Carlo experiments are called for in order to shed more light on the problem of selecting the most satisfactory optimal k-value for coping with the multicollinearity problem.

REFERENCES

Althauser, R.P. Multicollinearity and nonadditive regression models. In H.M. Blalock Jr. (ed.), Causal Models in the Social Sciences. Chicago: Aldine, 1971.

Blalock, H.M. Jr. Correlated independent variables: The problem of multicollinearity. Social Forces, 1963, 42, 233-237.

Blalock, H.M. Jr. Causal Inferences in Non-Experimental Research. Chapel Hill: University of North Carolina Press, 1964.

Bulcock, J.W., Lee, W.F. & Luk, W.S. The Theory of Normalization Ridge Regression. Report No. 2, Grant No. A3960, The Natural Sciences and Engineering Research Council of Canada, 1980. (Avail. ERIC).

Christ, C.F. Econometric Models and Methods. New York: John Wiley, 1966.

Dempster, A.P. Alternatives to least squares in multiple regression. In H.G. Kabe and R.P. Gupta (eds.), Multivariate Statistical Inference. Amsterdam: North Holland, 1973.

Farrar, D.E. & Glauber, R.R. Multicollinearity in regression analysis: The problem revisited. Review of Economics and Statistics, 1967, 92-107.

Gordon, R.A. Issues in multiple regression. American Journal of Sociology, 1968, 73, 592-616.

Hocking, R.R., Speed, F.M. & Lynn, M.J. A class of biased estimators in linear regression. Technometrics, 18, 1976, 425-437.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 1970a, 55-67.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Application to non-orthogonal problems. Technometrics, 12, 1970b, 69-82.

Hoerl, A.E. & Kennard, R.W. Ridge regression: Iterative estimation of the biasing parameter. Communications in Statistics, A5(1), 1976, 77-88.

Hoerl, A.E., Kennard, R.W. & Baldwin, K.F. Ridge regression: Some simulation. Communications in Statistics, 4(2), 1975, 105-123.

Intrilligator, M.D. Econometric Models, Techniques, and Applications, Englewood Cliffs, N.J.: Prentice Hall, 1978.

Johnston, J. Econometric Methods (2nd Edition). New York: McGraw-Hill, 1972.

Kasarda, J.D. & Shih, W.F. Optimal bias in ridge regression approaches to multicollinearity. Sociological Methods and Research, 5, 1977, 461-470.

Lawless, J.F. & Wang, P. A simulation study of ridge and other regression estimators. Communications in Statistics, A5(4), 1976, 306-323.

Leontif, W. Theoretical assumptions and nonobserved facts. American Economics Review, 61, 1971, 1-15.

McCabe, G.P. Jr. Evaluation of regression coefficient estimates using α-acceptability. Technometrics, 20, 1978, 131-139.

Marquardt, D.W. Generalized inverse, ridge regression, biased linear estimation, and non-linear estimation. Technometrics, 12, 1970, 591-612.

Marquardt, D.W. & Snee, R.D. Ridge regression in practice. The American Statistician, 29, 1975, 3-20.

Obenchain, R.L. Ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics, 17, 1975, 431-445.

Price, B. Ridge regression: Application to non-experimental data. Psychological Bulletin, 84, 1977, 759-766.

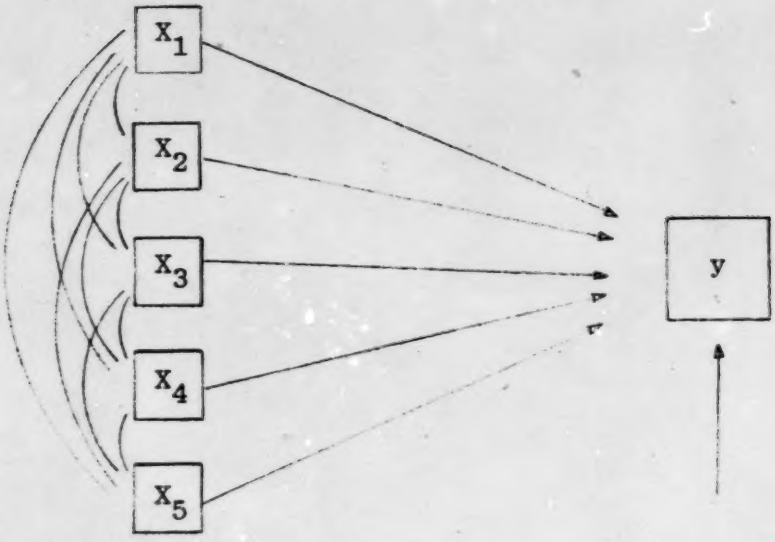
Vinod, H.D. Application of new ridge regression methods to a study of Bell system scale economics. Journal of the American Statistical Association, 71, 1976, 835-841.

Vinod, H.D. A survey of ridge regression and related techniques for improvements over ordinary least squares. Review of Economics and Statistics, 60, 1978, 121-131.

Wannacott, R.J. & Wannacott, J.H. Econometrics. New York: Wiley, 1969.

Wheeler, D. Human resource development and economic growth in developing countries: A simultaneous model. World Bank Staff Working Paper No. 407. Washington, D.C.: The World Bank, 1980.

Figure 1. Conceptual Model of the Five Factor Employee Satisfaction Study by Price (1977)[a]



a. Key to the mnemonics: How well the supervisor X_1 handled employee complaints; X_2 provided opportunities to learn new things about the job; X_3 gave raises based on performance; X_4 provided criticism of people who did a poor job; X_5 applied company policy fairly; X_6 the employee's assessment of the overall job being done by his/her supervisor. The sign on the paths represents the direction of the hypothesized relationship. u is the residual term.

Figure 2. Composite Variable Construction[a]

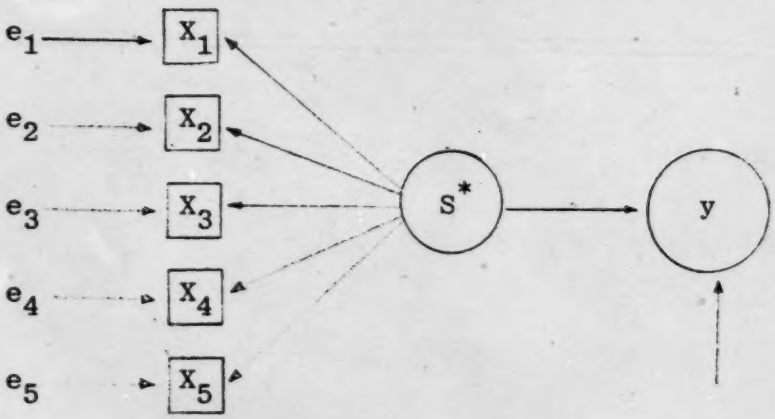


Fig. 2a. Example of Reflective Composite (Principal Component Model)

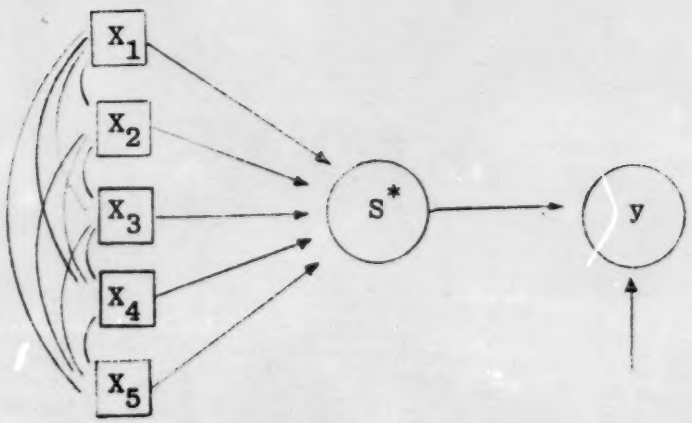


Fig. 2b. Example of Productive Composite (Canonical Correlation Model)

a. X_1 = COMPLAINTS, X_2 = LEARN, X_3 = RAISE, X_4 = CRITICISM, X_5 = POLICY, S^* = latent variable (Supervision Strategy), y = SATISFACTION.

Table 1. Correlation Matrix (X'X) and Eigenvalues for X'y for the Five-Factor Employee Satisfaction Study, Price, 1977

Variables	X_1	X_2	X_3	X_4	X_5	X_6
X_1 COMPLAINTS	1.000					
X_2 LEARN	.597	1.000				
X_3 RAISE	.669	.640	1.000			
X_4 CRITICISM	.188	.116	.377	1.000		
X_5 POLICY	.867	.819	.906	.207	1.000	
X_6 SATISFACTION	.827	.629	.592	.158	.763	1.000
λ_i [a]	3.348	.967	.401	.283	.001	

a. $\Sigma\lambda_i^{-1} = 738.7$

Table 2. VIF $(X'X)^{-1}$ Matrix of Predictor Variables for the Five-Factor Employee Satisfaction Study

Variables	X_1	X_2	X_3	X_4	X_5
X_1 COMPLAINTS	61.90				
X_2 LEARN	43.74	34.66			
X_3 RAISE	79.87	58.86	111.70		
X_4 CRITICISM	-13.91	-10.10	-19.53	4.47	
X_5 POLICY	-159.00	-117.50	-214.60	37.71	421.80

Vmax = 421.80, Dmax = .998, Dc = .999

Table 3. Performance Indices for OLS, RR, and NR Estimating Methods

Estimating Method	k	Vmax	Dmax	E.S.Ratio	RVAR	ϕ_0	α	Av.%Er.VAR	R^2
1. OLS	0	421.80	.998	1.00	0	0	1.000	12590.6%	.766
2. Normalzn.	.092	1.30	.186	1.41×10^{-6}	99.21%	26.02%	.295	33.0%	.636
3. K&S	.001	159.40	.996	.534	61.52%	3.65%	.967	956.8%	.729
4. HS&L	.053	1.59	.339	1.64×10^{-6}	99.02%	24.07%	.336	44.0%	.663
5. L&W	.061	1.50	.295	1.86×10^{-6}	99.07%	24.51%	.327	41.4%	.657
6. HK&B	.0013	129.9	.995	.487	68.44%	4.87%	.939	781.0%	.724
7. H&K Iter.	.141	1.07	.045	6.98×10^{-7}	99.36%	28.38%	.252	22.7%	.607
8. Vinod (#1)	.030	1.99	.498	2.50×10^{-6}	98.75%	22.48%	.373	11.7%	.681
9. Vinod (#2)	.964	.250	-.410	2.39×10^{-8}	99.87%	61.90%	.026	83.2%	.369

Table 4. Rankings of Performance Indices for RR and NR Methods

Performance Criteria	Normalzn	K&S	HS&L	L&W	HK&B	H&K Iter	Vinod #1	Vinod #2
k^a	6	1	4	5	2	7	3	8
$Vmax^a$	3	8	5	4	7	2	6	1
$Dmax^a$	3	8	5	4	7	2	6	1
$E.S.Ratio^b$	3	8	4	5	7	2	6	1
$RVAR^c$	3	8	5	4	7	2	6	1
$\phi_0{}^a$	6	1	4	5	2	7	3	8
α^d	Rej	Acc	Rej	Rej	Rej	Rej	Rej	Rej
Av.%Er.VAR	3	8	5	4	7	2	1	6
R^{2e}	6	1	4	5	2	7	3	8

a. Ranked from lowest (1) to highest (8)

b. Ranked from least sensitive to sampling error (1) to most sensitive (8)

c. Ranked from highest % reduction in variance (1) to lowest (8)

d. Whether the ridge estimate is α-acceptable (Acc) or not (Rej)

e. Ranked from highest R^2 (1) to lowest (8).